

Evolutionary isotonic separation for classification: theory and experiments

B. Malar · R. Nadarajan

Received: 26 January 2012 / Revised: 18 July 2012 / Accepted: 6 October 2012
© Springer-Verlag London 2012

Abstract Isotonic separation is a supervised machine learning technique where classification is represented as a linear programming problem (LPP) with an objective of minimizing the number of misclassifications. It is computationally expensive to solve the LPP using traditional methods when the dataset grows. Evolutionary isotonic separation (EIS), a hybrid classification algorithm, is introduced to tackle this issue. Here, isotonic separation acts as a host architecture where evolutionary framework based on genetic algorithm is embedded in the training phase of the isotonic separation, to find an optimum or near-optimum solution for the LPP. Evolutionary framework deploys a newly introduced slack vector to find the feasible solution. It also employs a position-based crossover operator to obtain the optimum or near-optimum solution. Experimental studies are conducted on Wisconsin Breast Cancer dataset and a synthetic dataset. Experimental and statistical results show that EIS outperforms its predecessors and state of the art machine learning techniques in terms of accuracy.

Keywords Isotonic separation · Evolutionary isotonic separation · Machine learning · Hybrid machine learning · Genetic algorithm

1 Introduction

Supervised machine learning attempts to estimate a function which maps instances from an input space \mathbf{X} to an output space \mathbf{Y} , given only a finite number of sample $\mathbf{D} = \{(x_i, y_i) \mid x_i \in \mathbb{R}^d, y_i \in \mathbb{R}\}_{i=1}^n$. Isotonic separation [4] is a supervised learning algorithm, which classifies data through an isotonic consistency condition, that is, if $i, j \in \mathbf{X}$ and $(i, j) \in R$ where R is a quasi relation if and only if $x_{ik} \geq x_{jk}$ for $k = 1, 2 \dots d$. These

B. Malar (✉) · R. Nadarajan
Department of Applied Mathematics and Computational Sciences, PSG College of Technology,
Coimbatore, Tamilnadu, India
e-mail: rsakthimalar@gmail.com

R. Nadarajan
e-mail: nadarajan_psg@yahoo.co.in

isotonic consistency conditions are derived from the domain knowledge of the problem. Isotonic separation formulates the classification problem as a linear programming problem with n decision variables and m constraints where $m = |R|$. Three major factors that influence the performance of the classifier/LPP are the number of instances n , the number of isotonic constraints m and the dimensionality d . In the worse case, a problem with n instances has $\frac{n(n-1)}{2}$ constraints. The complexity of the LPP is proportional to the size of the training set.

The limitation of traditional LPP solver [7] is that it needs to convert all inequality constraints into equality constraints using slack and surplus variables. This conversion increases the number of decision variables. Most of the optimization techniques start with an initial feasible point in the search space and move to the next point using some deterministic transition rule. For a large-scale LPP, getting an optimum solution is computationally expensive due to its time complexity $O((m+n)m^2)$. Even though the computing time of interior point method [22] is $O(\max(m, n)^3)$, it is infeasible due to excessive storage requirements and $m > n$ in isotonic separation.

The above linear programming problem can be viewed as a maximum flow network model with $n + 2$ nodes and $m + 2n$ arcs [4]. Even though efficient solution algorithms exist in the literature for maximum flow problems [1, 11], the computational time complexity of all algorithms is based on n and m . In isotonic separation $m \cong n^2$, it is computationally expensive to solve it for large values of n .

Recently, there has been an interest in the research community for biologically inspired systems in solving real-world optimization problems. Evolutionary algorithms (EA) are generalized optimization techniques inspired from natural evolution, which includes Darwin's survival of fitness principle [12]. An EA creates a pool of candidate solutions (chromosomes) called population that evolves over time and reaches optimal solution by means of genetic operators, such as mutation, crossover and selection. The typical applications of EA are difficult optimization problems for which specialized methods are not available or traditional methods fail for reasons whichever. Evolutionary techniques have widespread applications of machine learning such as instance selection [3], noise filtering [20] and data classification [10]. These motivated to use evolutionary approach for resolving the drawback of isotonic separation.

This paper proposes a hybrid model of isotonic separation, evolutionary isotonic separation (EIS) for classification. In EIS, isotonic separation acts as a host architecture where evolutionary framework is embedded in the training phase of the isotonic separation algorithm, to find an optimum solution for the large-scale LPP. This evolutionary framework deploys genetic algorithm (GA) to achieve this task. An extensive experimental evaluation is performed on the proposed hybrid model using Wisconsin Breast Cancer dataset (WBCD) and a synthetic dataset. For breast cancer diagnosis, experiments are conducted using isotonic separation, its variants and other state-of-the-art machine learning techniques. Experimental and statistical results show that EIS is more significant than the other algorithms for small- and large-scale classification.

The rest of the paper is structured as follows. Section 2 highlights the related literature for isotonic separation. Section 3 discusses a detailed description of the proposed EIS and its theoretical analysis. Section 4 demonstrates the proposed work with a case study. Section 5 reports the extensive experimental results of EIS for breast cancer diagnosis. It also reports the EIS experiments on a synthetic dataset. Section 6 concludes with a summary of the proposed framework and future enhancements.

2 Related work

Isotonic separation is a data classification technique, which has been applied in many fields such as internet content filtering [14], firm bankruptcy prediction [25] and breast cancer diagnosis [26].

Initially, isotonic separation method has been proposed for two-category separation and generalized to multicategory separation [4]. It aims to minimize the number of misclassifications by assigning penalty for type-1 and type-2 errors. If there are n instances of which n_1 and n_2 instances are misclassified as \mathbf{A}_1 and \mathbf{A}_0 , respectively, an objective function for the LPP is to minimize $\alpha n_1 + \beta n_2$, where α and β are penalties for misclassification. The relation R becomes the constraints of the LPP. A continuous outcome isotonic model has also been proposed. The major issues of isotonic separation are identified as feature selection and problem size reduction.

This method has been improved by grouping the instances with similar feature values [26]. Assume that there are a_i and b_i instances of \mathbf{A}_1 and \mathbf{A}_0 having the same feature values as x_i , the objective function of the LPP is

$$\min \alpha \sum_{i \in A_1} a_i (1 - y_i) + \beta \sum_{i \in A_0} b_i y_i \quad (1)$$

The constraints of the LPP are reduced by finding the maximal subsets \bar{A}_0 and \bar{A}_1 for classes A_0 and A_1 , respectively.

$$\begin{aligned} \bar{A}_0 &= \{i \in A_0 : \text{not} \exists j \in A_1 \text{ such that } (i, j) \in R\} \\ \bar{A}_1 &= \{i \in A_1 : \text{not} \exists j \in A_0 \text{ such that } (j, i) \in R\} \end{aligned} \quad (2)$$

Then, the class labels are assigned as follows:

$$y_i^* = \begin{cases} 1 & \text{if } i \in \bar{A}_1 \\ 0 & \text{if } i \in \bar{A}_0 \end{cases} \quad (3)$$

Now, the reduced subsets to be solved by LPP are

$$\begin{aligned} \mathbf{A}_0 &= \mathbf{A}_0 - \bar{\mathbf{A}}_0 \\ \mathbf{A}_1 &= \mathbf{A}_1 - \bar{\mathbf{A}}_1 \end{aligned} \quad (4)$$

This approach requires preprocessing tasks such as identifying similar data points in training. This method was trained and tested with 699 points and gave higher accuracy.

For firm bankruptcy prediction, isotonic separation method has been applied with a set of 23 financial ratios as features [25]. Isotonic separation has been tested and compared with nine other classification techniques. Even though it is a viable technique, the limitation in this scheme is that it has been trained and tested with small samples.

To filter the objectionable Internet content, isotonic separation has been deployed using Platform for Internet Content Selection (PICS) rating scheme [14]. A new approach has been proposed to include the test instances in a set of undominated points. The limitation of this approach is that it is computationally expensive to get the refined boundary points. Also, it requires the training set during testing, which consumes more memory, and it works like a lazy learner. This scheme has been trained and tested with a simulated dataset of 300 data points.

Luss et al. [18, 19] proposed a new technique, isotonic recursive partitioning, to address the major issues of isotonic regression, which is a nonparametric machine learning technique

for fitting monotonic models for data [24]. The problems of isotonic regression are statistical over-fitting and the computational difficulty due to higher dimensions. The basic concept behind isotonic recursive partitioning is to split the covariate space into smaller best cut problems. This creates a regularized sequence of isotonic models and converges to a global solution. A generalized Isotonic Recursive Partitioning is also proposed for fitting models for data [18, 19].

The potential weakness of the isotonic separation is classification of large dataset due to the complexity of solving LPP. EIS overcomes this drawback by embedding GA in the training phase of isotonic separation method. The main objective of EIS is to find an optimum or near-optimum solution to the large-scale LPP using evolutionary framework.

3 Evolutionary isotonic separation

The proposed EIS is a hybrid isotonic separation for classification, in which evolutionary framework is embedded in the training phase of isotonic separation to find the optimum solution of the LPP. This section gives the formulation of the mathematical model of isotonic separation into an optimization problem. It discusses a mathematical model for obtaining the optimum solution for the optimization problem using evolutionary framework. It also gives the stepwise procedure for hybrid classification algorithm, EIS.

3.1 Problem formulation

Given a finite set of objects \mathbf{A} , it is partitioned into two disjoint classes \mathbf{A}_0 and \mathbf{A}_1 , where $\mathbf{A} = \mathbf{A}_0 \cup \mathbf{A}_1$. Each object i has a d -dimensional feature space $x_i = (x_{i1}, x_{i2}, \dots, x_{id})$ where $x_i \in \mathbb{R}^d$ and a class label y_i where $i = 1, 2, \dots, n$

$$y_i = \begin{cases} 1 & \text{if } i \in \mathbf{A}_1 \\ 0 & \text{if } i \in \mathbf{A}_0 \end{cases} \quad (5)$$

The basic assumption in the isotonic separation [4] is isotonic consistency condition between objects. Based on this, a quasi ordering relation R is constructed.

$$R = \{(i, j) \mid i, j \in \mathbf{A} \text{ and } x_{ik} \geq x_{jk} \text{ where } k = 1, 2, \dots, d\} \quad (6)$$

The principal goal of the classifier is to develop a system that learns to classify the objects with minimum misclassifications. This is considered as an optimization problem, and the mathematical model is as follows:

$$\min \quad (7)$$

$$\alpha \sum_{i \in \mathbf{A}_1} (1 - y_i) + \beta \sum_{i \in \mathbf{A}_0} y_i$$

$$\text{subject to the constraints} \quad (8)$$

$$y_i - y_j \geq 0 \quad \text{for } (i, j) \in R$$

$$0 \leq y_i \leq 1 \quad \text{for } i \in \mathbf{A} \text{ (boundary constraint)}$$

Here, y_i is a binary variable, and it can be a real variable in $(0, 1)$. In the objective function of the LPP (7), $\sum_{i \in \mathbf{A}_1} 1 - y_i$ and $\sum_{i \in \mathbf{A}_0} y_i$ denote the number of type-1 and type-2 errors, respectively. Type-1 (type-2) error occurs when an object actually belongs to \mathbf{A}_1 (\mathbf{A}_0) and system misclassifies as \mathbf{A}_0 (\mathbf{A}_1). α and β are the penalties assigned for type-1 and type-2 errors. The main objective of the proposed work is to design an evolutionary framework that

finds an optimum solution for (7) so that the system developed from that optimum solution learns to classify unknown objects with minimum misclassifications.

3.2 Evolutionary framework

An evolutionary framework is a general purpose algorithmic framework that can be applied to different optimization problems with relatively few modifications. The optimization problem can be formulated as a model in order to apply GA [8, 12]. A model $M = (S, \Omega, f)$ of an optimization problem consists of:

- A search space S defined over a finite set of discrete decision variables $X_i, i = 1, 2, \dots, n$
- A set Ω of constraints among the variables $y_i - y_j \geq 0$ for $(i, j) \in R$
- An objective function $f(X) = \alpha \sum_{i \in A_1} (1 - y_i) + \beta \sum_{i \in A_0} y_i$

A feasible solution $s \in S$ is an assignment of values to variables that satisfies all constraints in Ω . A solution $s^* \in S$ is called a global optimum if and only if $f(s^*) \leq f(s) \forall s \in S$.

3.2.1 Chromosome representation

Evolutionary framework creates an initial population denoted by P , which is represented as $\mathbf{P} = \{C_i | C_i \in \mathbb{R}^n\}_{i=1}^N$ containing a set of N chromosomes. Each chromosome C_i is an n -dimensional vector: $C_i = (c_{i1}, c_{i2}, c_{i3}, \dots, c_{in})$ where c_{ij} represents a solution of the j th variable in the i th chromosome, $1 \leq i \leq N$ and $1 \leq j \leq n$, and n is the number of variables in the LPP, that is, $n = |\mathbf{A}_0 \cup \mathbf{A}_1|$. c_{ij} denotes the solution for the decision variable y_j , and it is restricted by a minimum and maximum $c_{\min} \leq c_{ij} \leq c_{\max}$. In general, initial population represents a potential feasible solution to the problem. But for the large-scale LPP, which has more constraints, obtaining the feasible solution as a chromosome is a challenging task. So this framework generates random chromosomes, aims to make these chromosomes to satisfy all the constraints, evaluates these feasible solution using fitness function and obtains the optimum or near-optimum solution with the help of crossover and mutation operators.

3.2.2 Fitness function

Isotonic separation generates a special type of LPP, in which all the constraints are in the form of $y_j - y_k \geq 0$, that is, the number of decision variables in each constraint is two and the right hand side of the constraint is always zero. The proposed framework takes this advantage and aims to find the feasible and optimum solution for (7). To achieve this goal, the framework introduces a slack vector for a population to check the feasibility of chromosomes. A slack vector is a one-dimensional array of size n , created for each chromosome while the constraints are checked. Based on the slack vector, the random chromosome is transformed into a feasible chromosome. This framework enables the reusability of a slack vector for the population.

Fitness function $f(\mathbf{C}) = f(f_1, f_2)$ is a function of two variables f_1 and f_2 , where f_1 is a function that maps a random chromosome to a feasible chromosome, and f_2 is a function that maps a feasible chromosome to a number of misclassifications. f_1 checks the random chromosome C_i with the constraints of the LPP and generates an n -dimensional slack vector $\Delta C_i = (\Delta c_{i1}, \Delta c_{i2}, \dots, \Delta c_{in})$ for chromosome. For every constraint in the LPP $y_j - y_k \geq 0 \forall k, k \neq j, 1 \leq k \leq n$, Δc_{ij} be the minimum slack value to be added to make c_{ij} as a feasible value for the decision variable y_j with respect to the set of constraints $y_j - y_k \geq 0$. This feasible value for a gene c_{ij} is obtained by taking the maximum of slack value Δc_{ij} for the respective gene in the random chromosome.

$$\Delta c_{ij} = \begin{cases} 0 & \text{if } c_{ij} - c_{ik} \geq 0 \\ \Delta c_{ij} & \text{if } \Delta c_{ij} > c_{ij} - c_{ik} \\ c_{ik} - c_{ij} & \text{otherwise} \end{cases} \quad (9)$$

$$c_{ij} = c_{ij} + \Delta c_{ij} \quad (10)$$

Algorithm 1 explains the procedure for converting a random chromosome into a feasible chromosome. Now, the feasible chromosome should be transformed to a binary chromosome by setting a threshold value θ , and normally $\theta = 0.5$. Then, the binary chromosome will be mapped into misclassifications using f_2 .

$$c_{ij} = \begin{cases} 0 & c_{ij} < \theta \\ 1 & \text{otherwise} \end{cases} \quad (11)$$

$$f_2(C_i) = \alpha \sum_{j \in A_1} (1 - c_{ij}) + \beta \sum_{j \in A_0} c_{ij} \quad (12)$$

Algorithm 1. Feasible Chromosome Conversion(C_i)

Input : Random Chromosome C_i & set of constraints

Output : Feasible Chromosome

While (all constraints are not satisfied)

For every constraint $y_j - y_k \geq 0$

If $c_{ij} - c_{ik} \geq 0$

$\Delta c_{ij} = 0$

Else

$diff = c_{ik} - c_{ij}$

If $diff \geq \Delta c_{ij}$

$\Delta c_{ij} = diff$

else

$\Delta c_{ij} = c_{ik} - c_{ij}$

End

$c_{ij} = c_{ij} + \Delta c_{ij}$

End

End

Return C_i

3.2.3 Elitism

From the population of chromosomes, the best chromosome with minimum number of misclassifications is taken to the next generation without applying genetic operators. This is done to preserve the best chromosome in the successive generations.

$$best = \arg \min_{C_i} (f_2(C_i)) \quad (13)$$

3.2.4 Crossover

The purpose of crossover operator is to generate a new chromosome that is different from their parents, yet containing some of their parent's features. To generate a new offspring, two parent chromosomes are selected using roulette-wheel selection, and then the values between the genes were swapped based on the crossover probability. A new approach is introduced to find the position of the genes to be swapped. The position of gene is chosen from the constraints $y_j - y_k \geq 0$ in the LPP, and the position vector P is created. Assume that C_1 and C_2 are two parent chromosomes selected for crossover, the positions are all j where j is in $y_j - y_k \geq 0$

$$P = \{j \mid y_j - y_k \geq 0 \forall k, k \neq j, 1 \leq k \leq n\} \quad (14)$$

$$c_{1j} = \begin{cases} c_{2j} & \text{if } r_{1j} \leq p_c \\ c_{1j} & \text{otherwise} \end{cases} \quad (15)$$

where j is the set of positions in the position vector P , r_{1j} is the random number generated for j th gene in the chromosome C_1 , and p_c is the crossover probability.

3.2.5 Mutation

It is one of the genetic operators to alter the values of one or more genes in the chromosome. This enables the GA to arrive at better solutions than the previous generation. For each individual gene in the chromosome, a random value r_{ij} is generated, and if it is less than the mutation probability p_m , then the gene value is inverted.

$$c_{ij} = \begin{cases} (1 - c_{ij}) & \text{if } r_{ij} \leq p_m \\ c_{ij} & \text{otherwise} \end{cases} \quad (16)$$

The brief procedure for the evolutionary framework is given in Algorithm 2.

Algorithm 2. Evolutionary Framework(Ω, f)

Input : Set of constraints Ω and the Objective function f

Output : Solution chromosome

1. For $i = 1$ to N
 - i. Generate C_i randomly
2. For each chromosome C_i
 - i. Call $C_i = \text{Feasible Chromosome Conversion}(C_i)$
 - ii. Obtain the binary chromosome using (10)
 - iii. Evaluate the fitness value using (11)
3. Preserve the best chromosome with minimum misclassification using (12)
4. For the rest of the chromosomes apply crossover using (13) and (14)
5. For the rest of the chromosomes apply mutation using (15)
6. Go to Step 2 and Repeat until there is no misclassification or the number of misclassifications reaches an acceptable threshold value.

3.3 Algorithm for evolutionary isotonic separation (EIS)

This subsection presents the stepwise procedure for classification using EIS based on evolutionary framework.

Step: 1. For the given training set \mathbf{A} , construct the optimization problem as given in (7) using the isotonic consistency condition (6).

Step: 2. Solve the optimization problem using EA as given in Algorithm 2, that is, $Y^* = \text{evolutionary framework}(\Omega, f)$. Here, Y^* be the optimum or near-optimum solution of the LPP returned from evolutionary framework.

Step: 3. Identify the boundary points or undominated points for each class using the below equation.

$$\begin{aligned} \mathbf{A}_1^* &= \left\{ i | y_i^* = 1 \text{ and } \exists y_j^* \in \theta^*, i \neq j, y_j^* = 1 \text{ and } (i, j) \in R \right\} \\ \mathbf{A}_0^* &= \left\{ i | y_i^* = 0 \text{ and } \exists y_j^* \in \theta^*, i \neq j, y_j^* = 0 \text{ and } (j, i) \in R \right\} \end{aligned} \quad (17)$$

where $\mathbf{A}_1^* \neq \emptyset$ and $\mathbf{A}_0^* \neq \emptyset$.

Step: 4. The d -dimensional space is separated into three regions based on undominated points:

$$\begin{aligned} Z_0 &= \{ j | (i, j) \in R \text{ where } y_i^* = 0 \} \\ Z_1 &= \{ j | (j, i) \in R \text{ where } y_i^* = 1 \} \\ Z_2 &= \mathbb{R}^d - (Z_0 \cup Z_1) \end{aligned} \quad (18)$$

Here, Z_0 denotes the region where all the points are labeled as 0, Z_1 denotes the region where all the points are labeled as 1, and Z_2 denotes the unclassified area where no data points exist.

Step: 5. In the test set, for every object h , whose attribute vector is $(x_{h1}, x_{h2}, \dots, x_{hd})$, classification is done as follows: If h lies in the area of Z_1 , then h belongs to class \mathbf{A}_1 . If h lies in the area of Z_0 , then h belongs to class \mathbf{A}_0 .

$$y_h^* = \begin{cases} 1 & \text{if } h \text{ lies in } Z_1 \\ 0 & \text{if } h \text{ lies in } Z_0 \end{cases} \quad (19)$$

If h lies in the area of where none of the objects exist and isotonic consistency condition cannot be able to determine its class label. In that scenario, the distance between the unknown object h and undominated points is measured. The object h is assigned to the class with minimum distance.

$$\begin{aligned} D_{h1} &= \beta \min \left\{ \sum_{k=1}^d \max(x_{hk} - x_{ik}, 0) \mid i \in Z_1 \right\} \\ D_{h0} &= \alpha \min \left\{ \sum_{k=1}^d \max(x_{ik} - x_{hk}, 0) \mid i \in Z_0 \right\} \\ y_h^* &= \begin{cases} 1 & \text{if } D_{h1} < D_{h0} \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (20)$$

3.4 Theoretical analysis of EIS

Lemma 1 *Slack vector always guarantees the decision variables to satisfy all the constraints of the LPP.*

Proof Let y_j be the decision variable in the LPP, and c_{ij} can be considered as a value for variable y_j and denotes the j th variable in the i th random chromosome. Assume that c_{ij} satisfies the boundary constraint. \square

Case: 1 *Only one constraint*

Let $y_j - y_k \geq 0$ be the constraint in the LPP. If the values of c_{ij} and c_{ik} satisfy the above constraint, then $\Delta c_{ij} = 0$ and c_{ij} becomes the feasible value for the decision variable y_j . Otherwise, $\Delta c_{ij} = c_{ik} - c_{ij}$ and $c_{ij} = c_{ij} + \Delta c_{ij}$. Substitute this updated value in the constraint

$$\begin{aligned} y_j - y_k &= c_{ij} - c_{ik} \\ &= c_{ij} + \Delta c_{ij} - c_{ik} \\ &= c_{ij} + (c_{ik} - c_{ij}) - c_{ik} \\ &= 0 \geq 0 \end{aligned} \quad (21)$$

In this scenario, slack vector alters the value of c_{ij} and makes it as feasible and the constraint is satisfied.

Case: 2 *More than one constraint*

Let $y_j - y_k \geq 0 \forall k, k \neq j, 1 \leq k \leq n$ be the set of constraints having y_j and it needs to be adjusted. If the value of c_{ij} is greater than all c_{ik} where $1 \leq k \leq n$, then all the constraints are satisfied $\Delta c_{jk} = 0$ and there is no change in the value of c_{ij} . Suppose, if c_{ij} is smaller than some c_{im} where $c_{im} \in c_{ik}$, then the constraints $c_{ij} - c_{im} \geq 0$ will not be satisfied by c_{ij} . In such a case

$$c_{im} = \{c_{ik} | 1 \leq k \leq n, k \neq j \text{ and } c_{ij} - c_{ik} < 0\} \quad (22)$$

Find the slack value for c_{ij}

$$\Delta c_{ij} = c_{im} - c_{ij} \quad (23)$$

The new updated value for c_{ij}

$$c_{ij} = c_{ij} + \Delta c_{ij} \quad (24)$$

Now the constraints that were not satisfied are

$$y_j - y_m \geq 0 \quad (25)$$

Substitute c_{ij} in (24) in the constraint (25),

$$\begin{aligned} y_j - y_m &= c_{ij} - c_{im} \\ &= c_{ij} + \Delta c_{ij} - c_{im} \geq 0 \end{aligned} \quad (26)$$

The combined value of $c_{ij} + \Delta c_{ij}$ is greater than c_{im} , and hence, the result of the constraint is definitely positive. Thus, slack vector always guarantees the decision variables to satisfy all the constraints of the LPP.

Lemma 2 *Given a chromosome and a set of constraints, the genes in the chromosome satisfy the boundary constraint.*

Proof Let y_j be the decision variable in the LPP, and c_{ij} can be considered as a j th gene in the i th random chromosome. $y_j - y_k \geq 0$ be the constraint in the LPP that contains y_j to be adjusted. Our assumption is that initial c_{ij} satisfies boundary constraint. \square

Case : 1 *only one constraint*

If the values of c_{ij} and c_{ik} satisfy the constraint, then $\Delta c_{ij} = 0$.

$$c_{ij} = c_{ij} + \Delta c_{ij} = c_{ij} + 0 = c_{ij} \quad (27)$$

There is no change in the value of c_{ij} , and it has already satisfied the boundary constraint. Thus, decision variable c_{ij} satisfy the boundary constraint.

If $y_{ij} < y_{ik}$, then the constraint is not satisfied $\Delta c_{ij} = c_{ik} - c_{ij}$ and $c_{ij} = c_{ij} + \Delta c_{ij}$. Substitute this updated value in the constraint

$$\begin{aligned} c_{ij} &= c_{ij} + \Delta c_{ij} \\ &= c_{ij} + c_{ik} - c_{ij} \\ &= c_{ik} \end{aligned} \quad (28)$$

Based on our assumption, c_{ik} satisfies the boundary constraint. So, new value of c_{ij} also satisfies the boundary constraint.

Case : 2 *More than one constraints*

Let $y_j - y_k \geq 0 \forall k, k \neq j, 1 \leq k \leq n$ be the set of constraints having y_j and it needs to be adjusted. If the value of c_{ij} is greater than all c_{ik} where $1 \leq k \leq n$, then all the constraints are satisfied $\Delta c_{jk} = 0$, and there is no change in the value of c_{ij} . Suppose, if c_{ij} is smaller than some c_{im} where $c_{im} \in c_{ik}$, then the constraints $y_j - y_m \geq 0$ will not be satisfied by c_{ij} . In such a case

$$c_{im} = \{c_{ik} | 1 \leq k \leq n, \quad k \neq j \text{ and } c_{ij} - c_{ik} < 0\} \quad (29)$$

$$\Delta c_{ij} = c_{im} - c_{ij} \quad (30)$$

The new updated value for c_{ij}

$$\begin{aligned} c_{ij} &= c_{ij} + \Delta c_{ij} \\ &= c_{ij} + c_{im} - c_{ij} \\ &= c_{im} \end{aligned} \quad (31)$$

$c_{im} \in c_{ik}$. Based on our assumption, c_{im} satisfies the boundary constraint. So, new value of c_{ij} also satisfies the boundary constraint.

Lemma 3 *Time complexity for an iteration of the evolutionary framework is $\theta(N(m + n))$.*

Proof Let m be the number of constraints and n be the number of variables in the LPP. Evolutionary model uses N chromosomes, and each chromosome has n real values. The computational time needed to generate a random chromosome is $\theta(n)$. To convert a random chromosome into a feasible chromosome and then to a binary chromosome, the time required is $\theta(m)$ and $\theta(n)$, respectively. Since there are N chromosomes, the running time required is $\theta(N(m + n))$. Fitness evaluation takes Nn computations. To find the best chromosome, number of comparisons required is $O(N)$. The time complexity to perform crossover and mutation in the population is $O(N)$. So the worst-case time complexity for the iteration of the evolutionary framework is $\theta(N(m + n))$. \square

Lemma 4 *Space complexity for an iteration of the evolutionary framework is $O(Nn)$.*

Proof Let m be the number of constraints and n be the number of variables in the LPP. Evolutionary model uses N chromosomes, and each chromosome has n real values. Total memory required to create a population is $O(Nn)$. During fitness evaluation, the memory required to create a slack vector is $O(n)$. \square

Theorem 1 *EIS terminates with an optimal solution for the LPP (7).*

Proof Let C_i be a chromosome in the population, and initially, it is a random, infeasible chromosome. A slack vector ΔC_i is generated by checking the constraints of the LPP. Feasible chromosome is obtained by adding the slack vector with chromosome. Based on Lemmas 1 and 2, it is shown that the feasible chromosome obtained from this process contains the feasible solution. For the feasible chromosome, fitness function is evaluated and the misclassifications are measured. The above process is done for all chromosomes, from which a chromosome with minimum misclassification is chosen as the best chromosome, and it is retained for the next generation. The remaining chromosomes will undergo mutation and crossover, and new offspring are generated. For the new offspring, the above procedure is done and the optimum solution is obtained. \square

Theorem 2 *Time complexity for evolutionary isotonic separation is $O(\max(Np(m + n), n^3))$.*

Proof Given a training set consists of objects in a d -dimensional data space, the time complexity to check the isotonic consistency constraints, and to construct a relation using (6) is $O(n^2d)$. If the relation R is represented with a graph data structure, the computational time needed to find out and eliminate transitive pairs is $\theta(n^3)$. Let m be the number of constraints after eliminating transitive pairs in R . From Lemma 3, it is shown that the computing time for an iteration of the evolutionary framework is $\theta(N(m + n))$. Let p be the number of generations for obtaining the optimum solution for the LPP in evolutionary framework. So, the time complexity of EIS is $O(\max(Np(m + n), n^3))$. \square

4 Case study

Consider a set U of 8 patients with two features, cell size and clump thickness, as shown in Table 1, in which patients 1, 4, 5 and 6 are diagnosed as malignant and the remaining patients are diagnosed as benign. The isotonic consistency condition is as follows: $P_{ik} \geq P_{jk}$ for $k = 1, 2$ where i and j are data points and k denotes feature-id. As a result of isotonic consistency condition, the following relation R is created.

$$R = \left\{ \begin{array}{l} (1, 2)(1, 3)(1, 4)(1, 5)(1, 7)(1, 8) \\ (3, 2)(4, 2)(4, 3)(4, 8)(5, 2)(5, 3) \\ (5, 4)(5, 7)(5, 8)(6, 2)(6, 3)(6, 4) \\ (6, 5)(6, 7)(6, 8)(7, 2)(7, 3)(7, 8) \\ (8, 2)(8, 3) \end{array} \right\}$$

With $\alpha = \beta > 0$, the LPP is formulated as follows:

$$\begin{array}{ll} \min & -y_1 + y_2 + y_3 - y_4 - y_5 - y_6 + y_7 + y_8 \\ \text{subject to} & \text{the constraints} \end{array}$$

Table 1 Data points

Id	Cell size	Thickness	Class
P1	8	10	1
P2	1	1	0
P3	2	1	0
P4	5	3	1
P5	8	7	1
P6	10	7	1
P7	6	1	0
P8	3	1	0

$$\begin{aligned}
 &y_1 - y_5 \geq 0 \quad y_3 - y_2 \geq 0 \\
 &y_4 - y_8 \geq 0 \quad y_5 - y_4 \geq 0 \\
 &y_5 - y_7 \geq 0 \quad y_6 - y_5 \geq 0 \\
 &y_7 - y_8 \geq 0 \quad y_8 - y_3 \geq 0 \\
 &\text{where } 0 \leq y_i \leq 1, 1 \leq i \leq 8
 \end{aligned}$$

To find a solution using evolutionary framework, initial random chromosome is generated with eight genes. Table 2 illustrates the process of converting a random chromosome into a feasible chromosome. The constraints are checked against the values of chromosome, and the slack values are calculated. If a variable has more than one slack value, maximum value is taken. When the constraints are checked with an initial chromosome, the first nonmatched constraint is $y_1 - y_5 \geq 0$. The slack value 0.2 is added to make c_{11} feasible. Similarly, all genes of the chromosome are updated to obtain the feasible chromosome. Table 3 demonstrates the evolutionary framework with a population of three chromosomes for two generations. As explained above, feasible chromosomes are obtained and converted into binary chromosomes by setting a threshold value 0.5. Then, fitness value is calculated by comparing a binary chromosome with actual class labels. The best chromosome is retained for the next generation without any modifications. The remaining chromosomes undergo crossover and mutation. The newly generated offspring are passed onto second generation. Here, if the convergence condition is assumed as number of misclassifications is one, this framework will come to an end after two generations and the values of third chromosome are treated as a near-optimum solution. Crossover is illustrated in Table 4 using two parent chromosomes of first generation from Table 3. Position vector $P = \{1, 3, 4, 5, 6, 7, 8\}$ is obtained from the constraints. There is no change in the first offspring due to elitism. To generate the second offspring, the first two chromosomes are assumed as parents. The italic value indicates that it is not involved in crossover. The bold value denotes that it has undergone crossover. After finding the optimum or near-optimum solution, boundary points are calculated for each class. Isotonic classifier finds the label of unknown instance h based on the distance between h and the boundary points.

5 Experimental analysis

To assess the effectiveness of the proposed EIS for breast cancer diagnosis, experiments are performed on Wisconsin Breast Cancer dataset using EIS and other state-of-the-art machine

Table 2 The evolution of an initial chromosome during the fitness evaluation

Iteration	Chromosome	Equation	c ₁₁	c ₁₂	c ₁₃	c ₁₄	c ₁₅	c ₁₆	c ₁₇	c ₁₈
1	Initial random	c_{ij}	0.3	0.5	0.2	0.9	0.5	0.7	0.6	0.8
	Genes to be checked	c_{ik}	0.5	–	0.5	0.8	0.9	0.6	0.5	0.8
	Slack value(s)	$c_{ik} - c_{ij}$	0.2	–	0.3	0	0.4	0.1	0	0.2
	Slack vector	Δc_{ij}	0.2	–	0.3	–	0.4	–	0.2	–
	Final	$c_{ij} = c_{ij} + \Delta c_{ij}$	0.5	0.5	0.5	0.9	0.9	0.7	0.8	0.8
2	Initial	$c_{ij} = c_{ij} + \Delta c_{ij}$	0.5	0.5	0.5	0.9	0.9	0.7	0.8	0.8
	Genes to be checked	c_{ik}	0.9	–	0.5	0.8	0.9	0.8	0.9	0.8
	Slack value(s)	$c_{ik} - c_{ij}$	0.4	–	0	0	0	0.2	0	0
	Slack vector	Δc_{ij}	0.4	–	0	0	0	0.2	0	0
	Final	$c_{ij} = c_{ij} + \Delta c_{ij}$	0.9	0.5	0.5	0.9	0.9	0.9	0.8	0.8
	Feasible	$c_{ij} = c_{ij} + \Delta c_{ij}$	0.9	0.5	0.5	0.9	0.9	0.9	0.8	0.8

Table 3 Example of an evolutionary framework with three chromosomes

Name	Iteration 1										Iteration 2						
Initial population	0.8	0.3	0.5	0.5	0.4	0.7	0.9	0.2	0.9	0.3	0.5	0.5	0.9	0.9	0.9	0.5	
	0.6	0.8	0.2	0.1	0.5	0.2	0.2	0.8	0.1	0.8	0.5	0.5	0.9	0.1	0.1	0.5	
	0.1	0.3	0.4	0.2	0.1	0.2	0.3	0.4	0.9	0.3	0.4	0.5	0.9	0.9	0.9	0.4	
Feasible population	0.9	0.3	0.5	0.5	0.9	0.9	0.9	0.5	0.9	0.3	0.5	0.5	0.9	0.9	0.9	0.5	
	0.6	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.9	0.8	0.8	0.5	0.9	0.9	0.5	0.8	
	0.4	0.3	0.4	0.4	0.4	0.4	0.4	0.4	0.9	0.3	0.4	0.5	0.9	0.9	0.9	0.4	
Binary population	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
	0	0	0	0	0	0	0	0	1	0	0	1	1	1	1	0	
Fitness value	3	4	4							3	4	4					
Elitised chromosome	0.8	0.3	0.5	0.5	0.4	0.7	0.9	0.5	0.9	0.3	0.4	0.5	0.9	0.9	0.9	0.4	
After crossover	0.9	0.3	0.5	0.5	0.9	0.9	0.9	0.5	0.9	0.3	0.5	0.5	0.4	0.3	0.5	0.9	
	0.9	0.8	0.5	0.5	0.9	0.9	0.9	0.5	0.9	0.8	0.4	0.5	0.9	0.9	0.9	0.4	
	0.1	0.3	0.4	0.5	0.1	0.9	0.9	0.4	0.9	0.3	0.4	0.5	0.9	0.9	0.9	0.4	
After mutation	0.9	0.3	0.5	0.5	0.9	0.9	0.9	0.5	0.1	0.7	0.5	0.5	0.6	0.7	0.5	0.1	
	0.1	0.8	0.5	0.5	0.9	0.1	0.1	0.5	0.9	0.2	0.6	0.5	0.1	0.9	0.9	0.4	
	0.9	0.3	0.4	0.5	0.9	0.9	0.9	0.4	0.9	0.3	0.4	0.5	0.9	0.9	0.9	0.4	

Table 4 The generation of an offspring after crossover and mutation

Parent-1	0.6	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
Parent-2	0.9	0.3	0.5	0.5	0.9	0.9	0.9	0.9	0.5
Offspring after crossover	0.9	0.8	0.5	0.5	0.9	0.9	0.9	0.9	0.5
Offspring after mutation	0.1	0.8	0.5	0.5	0.9	0.1	0.1		0.5

learning techniques, and the results are compared. In order to prove the effectiveness of EIS for large-scale dataset, a synthetic dataset is generated with nine features, and experiments are done on the data and compared with SVM and decision tree.

5.1 Dataset

The Wisconsin Breast Cancer dataset (WBC) [21] was obtained from the University of Wisconsin Hospitals, containing 699 data points taken from breast cancer patients. Among these, 458 and 241 are diagnosed as benign and malignant, respectively. The features used for classification are clump thickness, uniformity of cell size, uniformity of cell shape, marginal adhesion, epithelial cell size, bare nuclei, bland chromatin, normal nucleoli and mitoses. Each data point consists of values from 1 to 10, stating that a higher value corresponds to a more abnormal state of the tumor. In the case of breast cancer diagnosis, if a data point with certain values of clump thickness, uniformity of cell size, etc. is diagnosed as malignant, then other data points with the same or higher values must be diagnosed as malignant. Based on this, isotonic constraints are set for this domain.

In synthetic dataset, all the feature values are in the range of 0–10 in which lower value denotes the possibility of a malignancy due to that feature is less and higher value denotes more. Approximately 50 % of the data are in the range of 0–4, and the remaining data are in the range of 5–10. The class label of object i is calculated using maximal aggregation, in which maximum value among the features of a patient is considered [14]. The object is labeled using an independent Bernoulli random trial probability because it is empirically proved as a monotonically nondecreasing function on x_i in [14]. The probability is calculated as

$$p_i = \left(\frac{\text{sigmoid}(c(x_i - x)) - 0.5}{\text{sigmoid}(cx)0.5} \right) 0.5 + 0.5 \quad (32)$$

Where

$$x_i = \max(x_{ij}) \quad \text{where } 1 \leq j \leq d \quad (33)$$

$x = 5$ and c is a positive real number which controls the misclassification error, and sigmoid function is as follows:

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (34)$$

The class label y_i is determined by

$$y_i = \begin{cases} 1 & \text{if } p_i \geq 0.5 \\ 0 & \text{otherwise} \end{cases} \quad (35)$$

5.2 Experimental setup

Experiments are started with setting up an isotonic consistency condition. The partial ordering relation $R = \{(i, j) : P_{ik} \geq P_{jk} \forall k, k = 1, 2 \dots 9\}$ is created, and the reflexive and transitive pairs are eliminated. The resulting relation contains an average of 2,932 ordered pairs, and they become the constraints of the linear programming problem. Initial chromosomes should contain the feasible solution to the LPP. Here, initial chromosomes are generated randomly, and feasible chromosomes are obtained using a slack vector. The parameters and its values of the evolutionary framework are given in Table 5. These parameters are obtained by testing the framework for different values, and the values for which the framework gives better results

Table 5 Parameters of an evolutionary framework for WBCD and synthetic dataset

Parameter	WBCD	Synthetic
Population size	100	300
# Generations	920	2521
Elitism Rate	0.01	0.33
Crossover Rate	0.5	0.5
Mutation Rate	0.25	0.25
α	0.1	0.1
β	0.1	0.1

are taken. This model is tested by setting the convergence condition as either the number of correct classifications in the best chromosome is equivalent to the number of training instances or the number of correct classifications reaches an acceptable threshold value.

In Ryu's experiments, instances with identical feature values are combined together and the maximal subset is obtained. After preprocessing step, an LPP with 2,932 constraints and 630 variables is reduced to an LPP with 1,608 constraints and 265 variables. Jacob's isotonic separation experiments are replicated using tenfold cross-validation [14].

The proposed work is compared with state-of-the-art machine learning algorithms, SVM [5, 15, 27], decision tree and isotonic separation. SVM experiments are done using SVM light [16] with penalty parameter as 0.5 and linear kernel function. Decision tree [23] experiments are performed using Weka [13] classifier with a confidence factor of 0.25.

All experiments are done using tenfold cross-validation. Let \mathbf{A} be the given set of d -dimensional data for classification, which is partitioned into \mathbf{A}_i where $i = 1, 2, \dots, 10$ and each partition \mathbf{A}_i is stratified. In each trial i , partition \mathbf{A}_i is tested for accuracy measures using the model constructed with the training set $\mathbf{A} - \mathbf{A}_i$. Then, mean and standard deviation for the measures (accuracy, precision, recall, F-measure and receiver operating characteristic (ROC)) of 10 trials are calculated and reported [9].

In SVM [2], feature selection has been done with the help of a measure called F-Score and model parameters for RBF kernel functions are set using grid search. Nine models with different feature subsets based on F-score metric have been constructed and SVM experiments have been conducted. The accuracy of SVM is compared with the proposed EIS classifier. The best model comprises the five features including clump thickness, uniformity of cell size, uniformity of cell shape, bland chromatin and bare nucleoli. EIS experiments are done on three different length dataset (50–50 % training set and test set, 70–30 % training and test set, 80–20 % training and test set) with those features.

One-way ANOVA (analysis of variance) is done on the cross-validation results on each dataset to assess whether the mean test error rates of the EIS and state-of-the-art machine learning algorithms (SVM, decision tree and Isotonic separation) are different at the confidence level of 0.95. The p value denotes the probability under the null hypothesis that the mean error rates of all the algorithms are same. Smaller p value indicates the rejection of null hypothesis, which means that at least one algorithm is different from others. Then, t tests are conducted to assess the most significant algorithm against the others at the confidence level of 0.95 [6].

5.3 Results and discussion

Figure 1 shows the number of variables that are correctly classified in each generation and evolution of convergence toward zero misclassifications. It shows the results of convergence

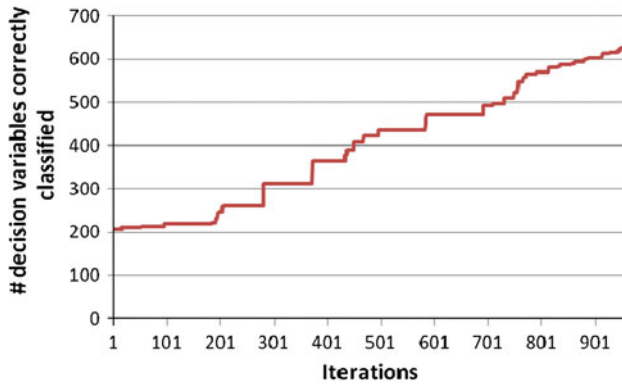


Fig. 1 Flow of convergence of evolutionary framework on Wisconsin Breast Cancer dataset

Table 6 Accuracy, precision, recall, ROC and F-measure values of EIS and other machine learning techniques on Wisconsin Breast Cancer dataset

	Accuracy (%)	Precision (%)	Recall (%)	TPR (%)	FPR (%)	F-measure (%)
Decision tree	92.4 ± 5.5	93.4 ± 4.3	93.0 ± 4.6	93.0 ± 4.6	8.2 ± 6.4	92.9 ± 4.8
SVM	96.6 ± 2.9	96.8 ± 2.8	96.5 ± 3.2	96.5 ± 3.2	3.4 ± 2.9	96.5 ± 3.1
IS	97.5 ± 2.6	96.1 ± 3.5	95.4 ± 3.8	93.4 ± 4.7	2.7 ± 3.8	94.7 ± 4.4
EIS	98.6 ± 0.7	97.5 ± 1.3	98.3 ± 1.2	99.5 ± 0.3	3.0 ± 1.1	97.6 ± 1.2

for the best model in evolutionary framework. For an LPP with 630 variables, GA converges in 920 iterations. The evolutionary framework enables to generate a chromosome with zero percentage misclassification. From the solution of the LPP, 67 and 43 boundary points are obtained for benign and malignant area. These points are used in classifying the test instances. When the dataset with nine features is tested using tenfold cross-validation, among 699 data points (among which 458 are benign and 241 are malignant), 428 data points belong to the benign area, 211 data points belong to the malignant area, and 60 data points belong to the unclassified area. Data points falling in the unclassified area are classified using the criterion. Among them, an average of 4 data points is misclassified.

The results of EIS are compared with its predecessors and the other classification algorithms and reported Table 6. Result show that evolutionary isotonic separation gives better result than the other algorithms in terms of accuracy, precision, recall, true positive rate, false positive rate and f-measure.

Three parameters are studied in the proposed work and the variants of isotonic separation: look ups, number of misclassifications in the LPP and accuracy of the classifier.

Look ups define the number of times the training set is accessed to construct a model during training and testing. In all variants of isotonic separation including EIS, generally training set is mandatory to check the isotonic consistency condition and to find the boundary points. In Ryu's isotonic separation, in addition to the above operations, training set is needed at the time of preprocessing to combine similar data points and to find the maximal set of data points to reduce the size of the LPP.

During testing, in all variants of isotonic separation except Jacob's technique, training set is not required to classify the data. Normally, a data point is classified by referring boundary points alone. But Jacob's technique repeatedly finds boundary points even after classifying a

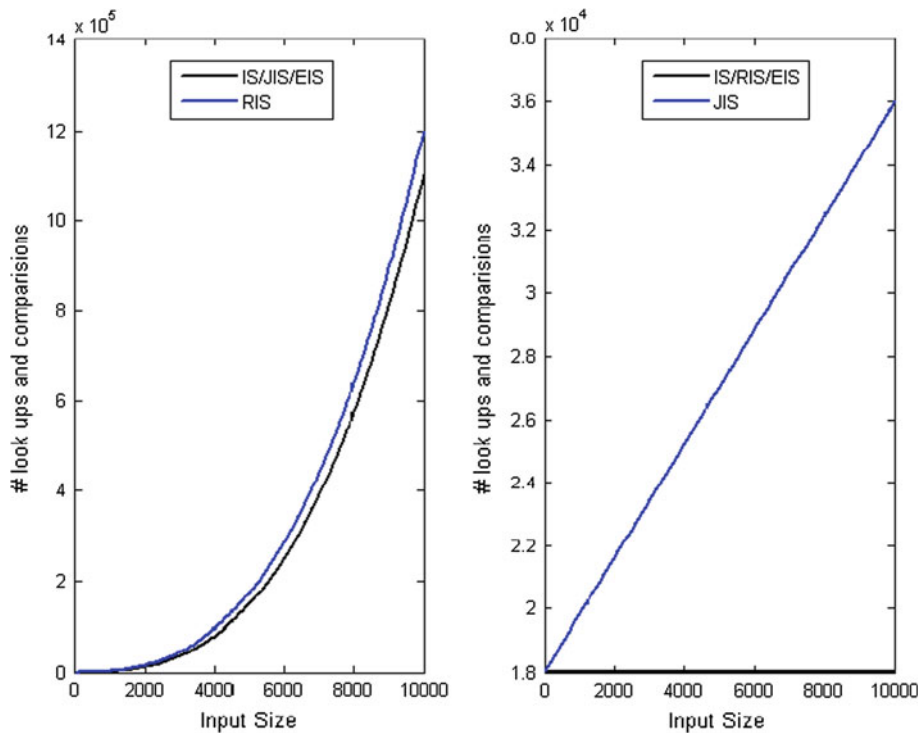


Fig. 2 Comparison of look ups of the training set in EIS and isotonic separation variants. (IS: isotonic separation [4] JIS: Jacob's isotonic separation [14] RIS: Ryu's isotonic separation [26] and EIS). (Left) X-axis represents size of training set and Y-axis represents the number of look ups and comparisons at training. (Right) X-axis represents size of test set and Y-axis represents the number of look ups and comparisons at testing

new data point. The unknown test instance is added to the set of boundary points by checking the isotonic consistency with the training set. So it requires that the training set should be available at the time of training. It is an overhead for the classifier at real time.

Figure 2a, b gives the number of look ups and comparisons required for different sizes of the training set during training and testing. During training, number of look ups is more for Ryu's technique due to preprocessing. All the other techniques require constant time complexity for training. EIS has less number of look ups; hence, it requires neither preprocessing as in Ryu's technique nor updation of boundary points while testing as in Jacob's technique.

The second parameter, the number of misclassifications in the LPP, plays a key role in the classification algorithm. From the solution of the LPP, the boundary points are measured for each class. In all the variants of isotonic separation, LPP is solved by using simplex or maximum flow network algorithm. Here, the solution obtained are real values and are interpreted as integer values. So, the misclassifications are more in jacob's method and in isotonic separation method. But Ryu's technique reduces the chance of misclassifications since it finds and eliminates the maximal subset of data points from the training set. In the proposed EIS technique, the evolutionary framework produces an optimum or near-optimum solution through an iterative process. In most of the times, the optimum solution would be a solution with zero misclassifications. Figure 3 gives the information about misclassifications in tenfold cross-validation experiments.

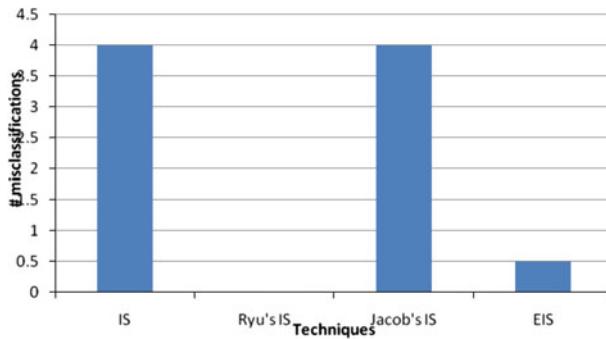


Fig. 3 Comparison of misclassifications occurred in the optimum solution of the LPP based on EIS and other isotonic separation techniques for Wisconsin Breast Cancer dataset X-axis denotes the techniques, and Y-axis denotes average number of misclassifications in the LPP during tenfold cross-validation

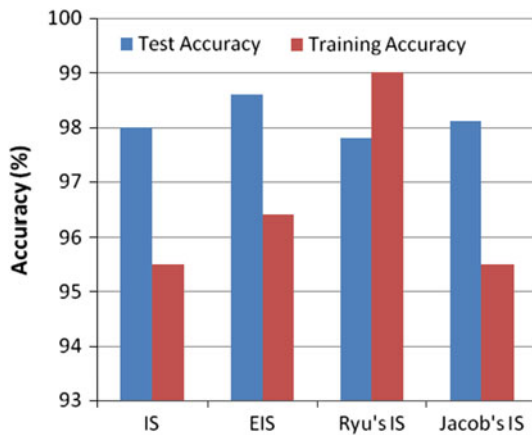


Fig. 4 Comparison of accuracy between EIS and other isotonic separation techniques for Wisconsin Breast Cancer dataset

The third parameter, accuracy of the classifier, determines how well the system classifies unseen data. The accuracy of the classifier is determined by the boundary points. Isotonic separation is able to generate more number of boundary points because it considers all instances in the training set. Since Ryu's technique eliminates maximal subset of data points from the training set for constructing LPP and finding boundary points, the number of boundary points obtained from it is less than isotonic separation. The accuracy measures of all the methods are projected in Fig. 4. Training accuracy is measured using the training set (trained and tested by training set), and test accuracy is measured with tenfold cross-validation. Results show that EIS gives higher accuracy than the other methods. The mean and standard deviation of precision, recall and F-measures of EIS are reported in Table 7. Since accuracy is not a correct measure at all times, receiver operating characteristic curve [9] is also reported. Since EIS is a discrete classifier, it has a point in the ROC space. Results show that EIS has marginally significant true positive rate than other methods, and it is not significant in reducing false-positive rate. From the results, it is shown that EIS gives better results than other methods in terms of precision, recall and F-measure.

Table 7 Comparative study of precision, recall, ROC and F-measure (in mean \pm standard deviation) of EIS and variants of isotonic separation

Method	Precision (%)	Recall (%)	F-measure (%)	TPR (%)	FPR (%)
Ryu's IS	93.6 \pm 4.5	91.1 \pm 5.8	89.8 \pm 7.1	95.2 \pm 3.6	1.4 \pm 2.3
Isotonic separation	96.3 \pm 3.4	96.1 \pm 3.1	95.5 \pm 3.8	94.9 \pm 1.2	1.4 \pm 1.9
Jacob's IS	96.8 \pm 1.2	91.4 \pm 2.1	93.2 \pm 2.4	95.2 \pm 2.1	1.8 \pm 2.6
EIS	97.5 \pm 1.3	98.3 \pm 1.2	97.6 \pm 1.2	97.1 \pm 1.9	1.9 \pm 2.1

Table 8 Accuracy, precision and recall measures of EIS and SVM with grid search and feature selection

Measures	Training set–Test set %					
	50–50 %		70–30 %		80–20 %	
	SVM	EIS	SVM	EIS	SVM	EIS
Malignant recall	99.55	93.41	99.24	95.24	100	97.65
Benign recall	96.64	98.98	98.61	99.75	97.91	100
Malignant Precision	98.22	98.40	99.24	90.91	98.88	100
Benign precision	99.14	96.03	98.61	98.32	100	99.44
Accuracy	98.53	96.94	99.02	98.86	99.51	98.96

Table 9 Statistical validation using one-way ANOVA on synthetic dataset

Source	Sum of squares	Degrees of freedom	Mean square error	F-value	Probability
Between	21.43	5	4.29	11.41	1.5e–7
Within	20.28	54	0.38		
Total	41.71	59			

Table 8 compares the performance of the proposed EIS with the grid-based SVM classifier. Results show that in 50–50 % training and test partition, more number of benign points and malignant points are correctly classified by the EIS than grid-based SVM. In 70–30 % partition of the training and test set, the number of benign points is more than SVM. In 80–20 % training and test partition, no malignant points are misclassified in EIS. But in SVM, no benign points are misclassified.

To prove the effectiveness of the proposed EIS statistically, one-way ANOVA followed by post hoc *t* test is conducted on tenfold cross-validation results of EIS and other techniques [6]. Table 9 shows that there is significant difference between the algorithms. Table 10 gives the results of *t* test to find the most significant algorithm at the confidence level of 95 %. Statistical results show that EIS has more significant than the other algorithms in providing lower test error rates at the confidence level of 95 %, even though it is not significant with Jacob's isotonic separation and SVM with feature selection using grid search.

For a synthetic dataset with 10,000 instances, the number of generations to obtain an optimum solution is 2,900. Using this solution, 194 and 186 points are identified as boundary points for malignant and benign regions, respectively. In SVM experiments, at the end of the training phase, 125 malignant and 129 benign points are identified as support vectors.

Table 10 Statistical validation using t test for EIS against other machine learning algorithms

Methods	Probability of t test
Isotonic separation	0.0001***
Ryu's isotonic separation	0.041**
Jacob's isotonic separation	0.91
SVM	0.05*
Decision tree	0.0009***
SVM with grid search	0.98

* Marginally significant ($p < 0.1$), ** Significant ($p < 0.05$), *** Very significant ($p < 0.01$)

Table 11 Comparative study of different measures of EIS and other machine learning techniques on synthetic dataset

Measures	Accuracy	Precision	Recall	F-Measure	TPR	FPR
Decision tree	96.3±1.2	96.5±1.8	87.7±3.5	87.0±1.5	93.2±2.6	1.8±2.1
SVM	95.9±2.3	92.3±3.5	75.6±3.7	83.1±3.6	91.7±3.2	0.5±1.6
Isotonic separation	NA	NA	NA	NA	NA	NA
EIS	97.9 ± 1.7	92.6 ± 3.4	98.8 ± 2.7	96.9 ± 3.9	96.8 ± 3.1	2.6 ± 1.8

Table 12 Statistical validation using t test for EIS against other machine learning algorithms on synthetic dataset

Methods	Probability of t test
SVM	0.04**
Decision Tree	0.02**

* Marginally significant ($p < 0.1$), ** Significant ($p < 0.05$), *** Very significant ($p < 0.01$)

The testing results of these methods are presented in Table 11. In case of synthetic data, experimental results show that EIS has higher accuracy (97.9%) than SVM (95.9%) and decision tree (96.3%). Table 12 presents the statistical validation through t test, and it shows that EIS is more significant than SVM and decision tree at the confidence level of 95%.

To compare the time complexity of evolutionary approach with other traditional methods to solve the LPP (Simplex, Interior point and Edmond–Karp max-flow algorithm), the parameters from the test instances of Wisconsin Breast Cancer dataset are taken and substituted in the upper bounds of time complexities. The time complexity of interior point method is $O(m^3)$, because in isotonic separation, $n \leq m \leq n^2$. Time complexity of Edmond–Karp algorithm to solve the max-flow network problem is $O(nm^2)$, which is computationally expensive. The comparative study of these methods is demonstrated in Fig. 5. From this, it is observed that, for both the datasets, EIS is an efficient than Simplex, Interior point or max-flow network models to obtain the optimum solution of the LPP. For large datasets, evolutionary approach is mandatory to train and classify the data using isotonic separation.

The only problem analyzed with EIS is that the convergence time has been increased with respect to the number of constraints and variables in the LPP. When the best chromosome achieves the near-optimum solution, it takes time to converge the optimum solution. To overcome this problem, a further investigation can be planned in future by combining the genetic operators with particle swarm optimization (PSO) [17].

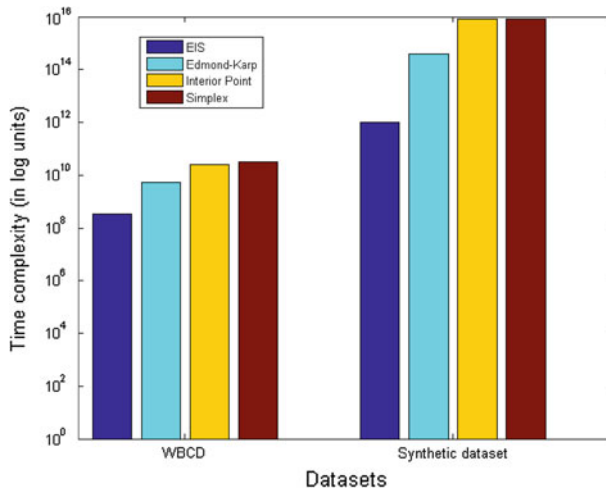


Fig. 5 Comparative study of time complexities between EIS, Interior point algorithm and Edmonds–Karp max-flow algorithms for the instances of WBC and synthetic datasets. X-axis represents the dataset, and Y-axis represents the time complexity of algorithms on dataset in logarithmic units

6 Conclusion

EIS is a hybrid isotonic separation method for classification in which evolutionary framework provides a solution for the computational difficulty of solving the large-scale LPP. EIS can be applied to problems where isotonic consistency condition exists between features and class label. This framework generates the random initial chromosome and converts it into feasible chromosome using a newly reusable slack vector. Experimental and statistical results signify that EIS resolves the drawback of isotonic separation, solving large scale LPP when the dataset grows, and it is proved to be a promising framework for large scale-classification.

Topics that remain to be explored in the future include evaluating the EIS in various domains such as Internet content filtering, mail spam filtering and other publicly available large datasets. In this paper, EIS is proposed for binary classification, and it can be generalized into multiway classification. In addition, finding the unknown isotonic consistency conditions for features constitutes a future research problem. Divide-and-conquer approach for implementing large-scale classification using isotonic separation can be analyzed and explored.

References

1. Ahuja RK, Magnanti TL, Orlin JB (1993) Network flows. Prentice-Hall, Englewood Cliffs
2. Akay MF (2009) Support Vector Machines combined with feature selection for breast cancer diagnosis. *Expert Syst Appl* 36:3240–3247
3. Cano JR, Herrera F, Lozano M (2003) Using evolutionary algorithms as instance selection For data reduction in KDD: an experimental study. *IEEE Trans Evolut Comput* 7(6):561–575
4. Chandrasekaran R, Ryu YU, Jacob V et al (2005) Isotonic separation. *INFORMS J Comput* 17(4):462–474
5. Christopher JCB (1998) A tutorial on support vector machines for pattern recognition. *J Data Min Knowl Discov* 2:121–167
6. Dawson RJM (1997) Turning the tables: A t-table for today. *J Stat Educ* 5(2)
7. Deb K (2001) Multiobjective optimization using evolutionary algorithm. Wiley, New York

8. Dorigo M, Birattari M, Stutzle T (2006) Ant Colony Optimization: Artificial Ants as a Computational Intelligence Technique. *IEEE Comput Intell Mag* 1(4): 28–39
9. Fawcett T (2006) An introduction to ROC analysis. *Pattern Recognit Lett* 27(8):861–874
10. Garcia-Piquer A, Fornells A, Orriols-puig A et al (2011) Data Classification through an evolutionary approach based on multiple criteria. *Knowl Inf Syst* doi:[10.1007/s10115-011-0462-9](https://doi.org/10.1007/s10115-011-0462-9)
11. Goldberg AV (1998) Recent developments in maximum flow algorithms. In: *Scandinavian workshop on algorithm theory*. Springer, London
12. Goldberg DE (1989) *Genetic algorithms for search, optimization and machine learning*. Addison-Wesley, Reading
13. Hall M, Holmes G, Pfahringer B et al (2009) The WEKA data mining software: an update. *SIGKDD Explor* 11(1):10–18
14. Jacob V, Krishnan R, Ryu YU (2007) Internet content filtering using isotonic separation on content category ratings. *ACM Trans Int Technol* 7(1):1–19
15. Joachims T (1998) Text categorization with support vector machines: learning with many relevant features. In: *European conference on machine learning (ECML)*
16. Joachims T (2002) SVM light support vector machine. <http://svmlight.joachims.org/>
17. Kennedy J, Eberhart R (1995) Particle swarm optimization. In: *IEEE international conference on neural networks*. Piscataway, pp 1942–1948
18. Luss R, Rosset S, Shahar M (2011a) Isotonic recursive partitioning. Submitted. Available at arXiv:1102.5496
19. Luss R, Rosset S, Shahar M (2011b) Generalised isotonic recursive partitioning. Submitted. Available at arXiv:1104.1779v1
20. Majid A, Lee CH, Mahmood MT et al (2011) Impulse noise filtering based on noise free pixels using genetic programming. *Knowl Inf Syst* doi:[10.1007/s10115-011-0456-7](https://doi.org/10.1007/s10115-011-0456-7)
21. Merz CJ, Murphy PM (1998) UCI repository of machine learning databases. Department of Information and Computer Sciences, University of California, Irvine
22. Monteiro R, Adler I (1989) Interior Path following primal-dual algorithms. Part II: convex quadratic programming. *Math Program* 44:43–66
23. Quinlan JR (1993) *C4.5: programs for machine learning*. Morgan Kaufman, San Mateo
24. Ryu YU, Chandrasekaran R, Jacob VS (2004) Prognosis using an isotonic prediction technique. *INFORMS J Manag Sci* 50(6):777–785
25. Ryu YU, Yue WT (2005) Firm bankruptcy prediction: experimental comparison of isotonic separation and other classification approaches. *IEEE Trans Syst Man Cybern Part A* 35(5):727–737
26. Ryu YU, Chandrasekaran R, Jacob VS (2007) Breast cancer detection using the isotonic separation technique. *Eur J Oper Res* 181:842–854
27. Watters CB, Shepherd M (2003) Support vector machines for text categorization. In: *Hawaii International conference on system sciences*. IEEE Computer Science Society

Author Biographies



B. Malar is an Assistant Professor in the Department of Applied Mathematics and Computational Sciences at PSG College of Technology at Coimbatore, India. She obtained her master's degree in Computer Applications and Master of Philosophy in Computer Science from Bharathiar University. She is pursuing her Ph.D. in Anna University, Coimbatore, India. Her research interests include machine learning, graph-theoretical approaches for machine learning and hybrid machine learning techniques using evolutionary and swarm intelligence.



R. Nadarajan received his master's and Ph.D. degrees in Mathematics in 1975 and 1981, respectively, from Annamalai University. He is presently the Professor and Head, Department of Applied Mathematics and Computational Sciences, PSG College of Technology, Coimbatore, India. His teaching and research interests span the fields of object computing, stochastic models, queuing theory, data mining, graph mining and context-aware computing. He has published over 70 research papers in refereed national and international journals. He has been the principal investigator for many research projects funded by government agencies. He is a member of ACM, Computer Society of India (CSI).